Vol.2012-IFAT-105 No.5
Vol.2012-NL-205 No.5
2012/1/20

Dual Decomposition for Chinese Semantic Role Labeling

Yanyan Luo,[†1] Masayuki Asahara [†1]
and Yuji Matsumoto [†1]

Semantic role Labeling(SRL) is gaining more attention as it benefits a wide range of natural language processing applications. In the past few years, dependency-based SRL has attracted much attention. However, the bottleneck for Chinese dependency parsing severely limits the ability to solve Chinese SRL. Chinese shallow parsing has better results, but only using this kind of features cannot get ideal results for SRL over word units. This paper describes a dual decomposition method to get the final inference, which can leverage both the dependency parsing information and shallow parsing information. In the experiments, the proposed model achieved much performance improvements.

1. Introduction

We investigate the possibility to construct an effective joint system for Chinese SRL. This single joint system is built by embedding two independent SRL models: dependency-based SRL and shallow parsing-based SRL. The full model allows the two embedded models to disagree to some extent, but to reward agreement.

Dependency-based SRL is the task to assign argument role labels to nodes of dependency parsed trees using dependency information. Dependency-based SRL was first explored by Hacioglu[1] and was promoted by the CoNLL-2009 shared task[2]. In this challenge, predicate sense disambiguation was also considered and better understanding of SRL with dependency parsing is achieved. The task showed, SRL can obtain better performance with better dependency parsing results. Chinese dependency parsing faces some serious bottlenecks that severely limit the ability to solve Chinese SRL. Whereas, Chinese shallow parsing has gained a promising result[3] which indicates an alternative choice for Chinese SRL. Traditionally, shallow parsing based SRL systems are implemented over chunk units or constitute units. The difference between a chunk and a constitute is that the latter is comprised of one or more chunks. Although these systems achieved promising results, for example Sun's work in 4). It is still unknown how the SRL performs directly over word units using shallow parsing based features.

The paper addresses the Chinese SRL problem over words of a sentence on the basis of shallow syntactic information at the level of phrase chunks. Meanwhile, we employ dual decomposition[5] method as an approximate inference technique to joint dependency-based Chinese SRL and shallow parsing-based Chinese SRL. In this approach, we make the two models work together and get more improvements in the final result.

The rest of the paper is organized as follows: related work about our research is listed in Section 2. Section 3 and Section 4 outline the dependency-based Chinese SRL and shallow parsing-based Chinese SRL respectively. In Section 5 we describe the dual decomposition method to achieve the final SRL result. Section 6 shows the evaluation results and finally the conclusion and future work are listed in Section 7.

2. Related Work

Most existing systems for automatic Chinese SRL make use of the full constituent parse of a sentence to define argument boundaries and to extract relevant information for training classifiers. On the contrary, Sun[4] addresses Chinese SRL problem on the ba-

†1 Nara Institute of Science and Technology

sis of shallow syntactic information at the level of phrase chunks. In their approach, Chinese SRL is formulated as a sequence labeling problem, performing IOB2 decisions on the syntactic chunks of the sentence. However, this method is useless for SRL over word units since a chunk consists of successive words. Little is known about how shallow parsing-based SRL performs over word units.

A substantial amount of research has focused on dependency-based SRL since CoNLL-2009 shared task and some progressing achieved in jointly implementing predicate sense disambiguation and SRL. Watanabe[6)] proposes using dependencies between predicate senses and semantic roles to implement predicate sense disambiguation and SRL simultaneously and proves that both tasks can help each other. Unfortunately, the system is designed primarily depending on the performance of the dependency parsing, it is difficult to implement a state-of-art SRL system without good dependency parsing results.

Recently, dual decomposition has become a hot framework for joint model. It has been successfully used in non-projective dependency parsing[7)], word alignment[8)], etc. Dual decomposition builds a connection to linear programming which ensures that the algorithms provide a certification of optimality when they recover the exact solution. Inspired by those work, we construct a SRL model by integrating dependency-based SRL and shallow parsing-based SRL with an efficient and easy implementation and this model can alleviate the heavy dependence on dependency parsing.

3. Dependency-based Chinese SRL

Our dependency-based SRL model consists of two tasks: predicate sense disambiguation and argument classification. Both tasks are jointly implemented which alleviates the error propagation problem in pipeline methods and makes the two tasks to be able to help each other.

3.1 Model Definition

To obtain a probabilistic model for complete label sequence of arguments and the given predicate, $Pr(S|X)$, we define the models for predicate sense disambiguation $Pr(P|X,p)$ and arguments classification $Pr(A|X,P,p)$. Then we perform the following probabilistic function for label assignments:

$$Pr(S|X) = Pr(P|X,p) \times Pr(A|X,P,p) \qquad (1)$$

where X is a word sequence with P denoting the sense of the given predicate word in location p; A is an argument assignment sequence and S is a sequence that consists of A and P. In order to capture global clues of SRL, we use exponential models as follows:

$$Pr_{global}(S|X) = \frac{1}{Z(X)} Pr(S|X) exp \left\{ \sum_{m=1}^{M} \gamma_m f_m(S,X) \right\} \qquad (2)$$

where $f_m(S,X)$ is the m-th global feature function, M is the number of global feature functions and γ_m is the weight of the m-th feature. To get the normalization factor $Z(X)$ over the whole sequences of S, we need to perform computationally expensive search. As done in previous work[6)], we use a simple approach, N-best relaxation.

3.2 Features for Dependency-based SRL

A wide range of features have been shown to be useful in previous work on SRL[9)]. In our experiments many features that described in recent work on dependency-based SRL are adapted to Chinese. We briefly discuss these features and divide these features into two sets: local features and global features.

3.2.1 Local Features

- Path: lemma, POS and dependency paths between the argument candidate in the focus word and the predicate.
- Position: The position is defined in relation to the predicate and the values are "before", "after" and "equal".
- Family: The position of the argument candidate with respect to the predicate position in the dependency tree, such as "child", "parent", "grandchild".
- Argument candidate word features: Lemma and POS of the focus argument candidate, its parent, leftmost/rightmost child, leftmost/rightmost sibling.
- Predicate word features: Lemma and POS of the predicate, predicate's head.
- Predicate dependency: The dependency label between the predicate and the predicate's head.
- Predicate classes: Different with the classification defined in Xue[10], we donot classify the predicate along three dimensions but two dimensions⋆1 : the number of framesets, the number of arguments under a special frameset. A frameset roughly corresponds to a major sense. For example, a predicate may be presented as $C1C2$, which means that the predicate has two main meanings, with the first sense having one argument and the second having two arguments.
- Dependency: Dependency labels of the focus argument candidate and its dependents.
- Pair-wise features: These features consider both the sense of the predicate and argument label assignment simultaneously, In our system, the lemma of the current argument candidate; POS of the current argument candidate; combination of its lemma and POS; dependency label path between the argument candidate and the predicate in the dependency parsed tree are used. Unlike the predicted predicate sense in Johansson[9], we also use other predicate sense candidates, which makes the argument classification results to help predicate sense disambiguation synchronously and vice versa.

3.2.2 Global Features

- Predicate-Argument label sequence: The sequence of the predicate sense and argument labels, e.g. $A_0 - pred.sense - A_1$.

4. Shallow Parsing-based Chinese SRL

4.1 Chinese Shallow Parsing

Although the research on Chinese chunking has been done for many years and a variety of chunk definitions have been proposed, most of them did not provide sufficient detail. In our system, we use chunk definition presented in Chen[3]. The definition of syntactic chunks is illustrated in line CH in **Fig. 1**.

WORD	去年 last year	西藏 Tibet	金融 finance	工作 work	取得 get	显著 remarkable	成绩 achievement
POS	NN	NR	NN	NN	VV	JJ	NN
CH	[NP]	[NP]	[NP	]	[VP]	[ADJP]	[NP]
SRL	TMP	NONE	NONE	A0	取得.01	NONE	A1

"The finance work in Tibet got remarkable achievement last year."

Fig. 1 An example of Chinese SRL over words .

With the IOB representation, the problem of Chinese chunking can be regarded as a sequence labeling task. In our system, we

⋆1 This rough classification can be automatically derived from the frame files, which are created to guide the Chinese Propbank annotation.

use the CRF++ tool[⋆2] with the following feature templates to perform $IOB2$ decisions on the chunks of a sentence:

- Uni-gram word/POS tag features:x_{-2}, x_{-1}, x_0, x_1, x_2.
- Bi-gram word/POS tag features:$x_{-2}x_{-1}$, $x_{-1}x_0$, x_0x_1, x_1x_2.

4.2 SRL with Shallow Parsing

The assignments of argument labels and predicate sense label are illustrated in Fig. 1. In this example, the head word ”工作” of the noun phrase ”金融工作” (finance work) is labeled as A_0, meaning that it is an *Agent* of current predicate ”取得” (obtains).

As for the model definition, the model defined in Section 3.1 can still be adopted. Obviously, there are two kinds of words in a sentence, i.e. head words and non-head words. Besides the global and local division metric, the local features can be subdivided into features for head words and for non-head words.

4.3 Local Features

For non-head words, the token-level features used in the system just include: lemma and POS tags. No matter whether the predicate word is a head word of a chunk, the POS tag of the predicate; the POS tag of the words that immediately precede and follow the predicate; the lemma of the predicate; the lemma of the words that immediately precede and follow the predicate and the predicate classes defined above are applied.

For the head word of a chunk the following local features are defined:

- Chunk Type
- Length: the number of words in a chunk.
- Head word/POS tag: we use the rules in Jurafsky[11] to extract head words.
- Chunk word/POS tag context: chunk context includes one word before and one word after the chunk.
- Position: the position of the chunk with respect to the predicate.
- POS tag sequence: the POS tag sequence of words in a chunk. For example, ”金融工作” (finance work) is NN_NN.
- IOB chunk tag of head word: chunk tag of head word with IOB representation, e.g. B-NP.
- Predicate lemma/POS tag context: the lemmas /POS tags of the words that immediately precede and follow the predicate.
- POS chain: includes the POS tags of head words that between the token and the predicate.
- Chunk number: the number of chunks between the head word and the predicate.
- Combined features: conjunctions of position and head word; position and predicate word; position, head word and predicate word; position and predicate classes; position and POS tag of head word; position and POS tag of predicate; position, predicate classes and head word; position, predicate classes and POS tag of the head word.
- Pair-wise features: the same functions with those in dependency-based SRL, these features both consider the sense of the predicate and argument label assignments simultaneously. The POS tag of the argument candidate (head word); the combination of the lemma and POS tag of the argument candidate; the lemma of the argument candidate are used in our system.
- Verb formation: these features proposed by Sun[4] and a verb formation analyzing method is given.

⋆2 http://crfpp.sourceforge.net

Vol.2012-IFAT-105 No.5
Vol.2012-NL-205 No.5
2012/1/20

5. Integration of Two Chinese SRL Models

This section describes the dual decomposition approach for the final inference of Chinese SRL.

5.1 Dual Decomposition Algorithm

Dual decomposition is a classical method for solving optimization global problems that can be decomposed into efficiently solvable sub-problems. According to Rush[5], the dual decomposition algorithm has the following features:

- Simple: it uses basic dynamic programming algorithms.
- Efficient: it is faster than full dynamic programming intersections.
- Strong Guarantees: this algorithm has strong theoretical guarantees in guaranteed convergence and the possibility of a certificate of optimality. These guarantees are possible because the algorithms directly solve an LP relaxation.

5.2 Dual Problem Formulation

The goal is to get better SRL performance without greatly increasing the complexity of inference. First, we define the predicate-argument label sequences from dependency-based SRL and shallow parsing-based SRL are S_D and S_H respectively. Meanwhile, their corresponding sequence sets are $S_D(X)$ and $S_H(X)$. Also, we defined $Pr_{global}(X, S_D)$ to be the normalized probability of an assignment of S_D. Similarly, we defined $Pr_{global}(X, S_H)$ to be the normalized probability of an assignment of S_H. Finally, we define the index set for SRL to be $\mathbb{I} = \{(i,t) : t \in T(P) \text{ when } i = p; t \in T(A) \text{ when } i \neq p\}$. As defined in Section 3.1, p still indexes the location of the predicate in an instance X and i indicates the location of the argument in the instance. $T(A)$ and $T(P)$ are argument label set and predicate sense set respectively. The integrated problem is then to solve

$$\begin{aligned} &\underset{S_D, S_H}{argmax}\, Pr_{global}(X, S_D) + Pr_{global}(X, S_H) \\ &\textit{such that: } S_D(i,t) = S_H(i,t) \quad \forall (i,t) \in \mathbb{I} \end{aligned} \tag{3}$$

5.3 Training

We choose prediction-based Passive-Agrressive(PA) online learning[12] with parameter averaging technique[6] to estimate the weights. PA is an error-driven learner that shifts weights towards features of the gold solution and away from features of the current guess, whenever the current model makes a mistake. PA learning takes into account a user-defined loss function for which we defined as the number of incorrect assignments for arguments and predicate sense.

6. Experiments and Discussion

6.1 Experimental Settings

We perform experiments over the CoNLL-2009 shared task[2] dataset for Chinese. In this dataset, there are 8,104 predicates in development corpus; 102,810 predicates in training corpus and 10,282 predicates in test corpus.

The number of iterations for PA algorithm was set to 5. Since it difficult to calculate all the possible assignment sequences for global model, we generated N-bests by local model to apply global features. A large N makes the training and calculation expensive and more N local outputs would lead to more local results having the same global features. In our experiments, the value of N was 3. The evaluation software provided by CoNLL-2009 shared task defined the following final evaluation criteria and the same with the task, we assumed the predicates have been identified.

$$Precision(P) = \frac{\#of\ correctpred.senses + \#of\ correctarg.roles}{\#predicates + \#ofreturnedarg.roles}$$

$$Recall(R) = \frac{\#of\ correctpred.senses + \#of\ correctarg.roles}{\#predicates + \#arguments}$$

$$F_1 = \frac{2PR}{P + R}$$

6.2 Results and Discussion

Table 1 [*3] shows the performances of SRL in our model. Comparing the first two lines, it can be seen that shallow parsing-based SRL performances worse in both recall and precision and especially in recall. When we introduce dual decomposition method, the performance of SRL is greatly improved to 79.59. These results suggest that even good labeling performance has been achieved by dependency based SRL, partial parse based SRL can still enhance their performance.

6.3 SRL Performance with dual decomposition

In order to see how much improvements can be obtained from dual decomposition method, **Fig. 2** shows the changes of the SRL performance with different maximum iterations on development corpus. Although there is a little fluctuation that the F value degrades from 90.58 to 90.54, when the iterations becomes from 10 to 20, it still can be seen that SRL can be benefited from the dual decomposition.

Table 1 Results on Chinese test dataset of CoNLL-2009 shared task

Systems	P	R	F_1
Dependency-based SRL	82.94	75.36	78.97
Shallow parsing-based SRL	80.77	66.62	73.02
SRL with dual decomposition	84.52	75.2	79.59

*3 The value of maximum iteration times in dual decomposition equals 5 .

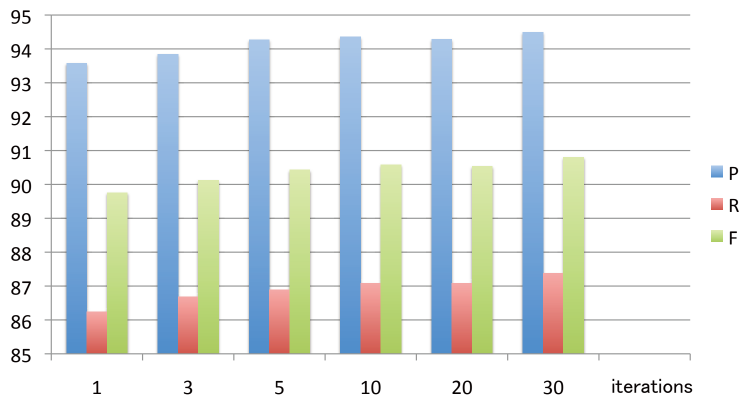

Fig. 2 Performance of SRL assuming a fixed number of iteration on development corpus.

7. Conclusion and Future Work

We implemented dependency parsing-based SRL and shallow parsing-based SRL respectively. Although dependency-based SRL has better performance than chunking-based SRL, our experiments showed that shallow parsing SRL can still helpful for SRL. By explicitly capturing constraints between both systems, we achieved great improvements for SRL. We used a dual decomposition method and supervised online learning, this recipe can be successful in many settings. Although the runtime penalty is kept minimal by using dual decomposition, since the expensive computation time is required for independent SRL systems, the total time consuming is still considerable. **Table 2** shows the asymptotic of our models with respect to one sentence with m predicates; n candidate arguments; each predicate has $|T(P)|$ senses; $|T(A)|$ semantic role labels in the corpus and R iterations

Vol.2012-IFAT-105 No.5
Vol.2012-NL-205 No.5
2012/1/20

Table 2 Time complexity analysis

Systems	Complexity
Dependency-based SRL	O(m$\|T(P)\|$n$\|T(A)\|$)
Shallow parsing-based SRL	O(m$\|T(P)\|$n$\|T(A)\|$
SRL with dual decomposition	O(Rm$\|T(P)\|$n$\|T(A)\|$

for dual decomposition. In the future, we want to apply some pruning technique to reduce the number of argument candidates and apply this method to English SRL.

References

1) Hacioglu, K.: Semantic Role Labeling Using Dependency Trees, *Proceedings of the 20th international conference on Computational Linguistics*, pp.1273–1276, Montreal, Quebec, Canada, 2004. Association for Computational Linguistics.
2) Hajič, J. et al: The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languagess, *Proceedings of the 13th conference on Computational Natural Language Learning*, pp.1–18, Boulder, Colorado, 2009. Association for Computational Linguistics.
3) Chen, W.L . et al: An Empirical Study of Chinese Chunking, *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pp.97–104, Sydney, Australia, 2006. Association for Computational Linguistics.
4) Sun, W.W. et al: Chinese Semantic Role Labeling with Shallow Parsing, *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp.1475–1483, Singapore, 2009. Association for Computational Linguistics.
5) Rush, M. et al: On Dual Decomposition and Linear Programming Relaxations for Natural Language Processing, *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp.1–11, Cambridge, Massachusetts, 2010. Association for Computational Linguistics.
6) Watanabe, Y. et al: Multilingual Syntactic-Semantic Dependency Parsing with Three-stage Approximate Max-margin Linear Models, *Proceedings of the 13th conference on Computational Natural Language Learning*, pp.114–119, Boulder, Colorado, 2009. Association for Computational Linguistics.
7) Koo, T. et al: Dual Decomposition for Parsing with Non-projective Head Automata, *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp.1288–1298, Cambridge, Massachusetts, 2010. Association for Computational Linguistics.
8) DeNero, J. and Macherey, K: Model-based Aligner Combination Using Dual Decomposition.*Proceeding of the 49th Annual Meeting on Association for Computational Linguistics*, pp.420–429, Portland, Oregon, 2011. Association for Computational Linguistics.
9) Johansson, R. and Nugues, P: Dependency-based Semantic Role Labeling of PropBank. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp.69–78, Honolulu, 2008. Association for Computational Linguistics.
10) Xue N.W.: Labeling Chinese Predicates with Semantic Roles. *Computational Linguistics*, Vol.34, No.2, pp.225–255(2008).
11) Sun H.L and Jurafsky D.: Shallow Semantic Parsing of Chinese. *HLT-NAACL*, pp.233–240, Boston, 2004. Association for Computational Linguistics.
12) Crammer K.: Online Passive-aggressive Algorithms. *Journal of Machine Learning Research*, Vol.7, pp.551–585(2006).